2022 International Symposium on Nonlinear Theory and Its Applications,
NOLTA2022, Virtual, December 12-15, 2022

Time Series Classification by Neural Network Using Features of Attractors after Smoothing Process

Ryosuke Shimizu†, Yoko Uwate‡, and Yoshifumi Nishio⋆

†Dept. of Electrical and Electronic Engineering, Tokushima University
2-1 Minami-Josanjima, Tokushima 770–8506, Japan
Email: shimizu@ee.tokushima-u.ac.jp, uwate@ee.tokushima-u.ac.jp, nishio@ee.tokushima-u.ac.jp

Abstract—Time series classification is an important and challenging problem in data analysis. Recently, time series analysis using neural networks (NN) has attracted much attention. However, the analysis of time series data with complex oscillations is difficult. Therefore, it is important to search for effective features of the data. In this study, we transform the dimensionality of the data and search for features suitable for NN classification. In this study, we investigate the effect of smoothed data on the classification accuracy of the dimensionality reduction method and the features of the data.

1. Introduction

In recent years, there has been a great deal of research on the analysis of time series data. This real-world data has various characteristics, such as not only periodic oscillation but also random oscillation. Among them, chaos theory deals with data that obey deterministic laws. However, in order to observe chaos, it is necessary to transform the data. One of the methods is the time-delay coordinate system, which is a kind of chaos theory [1]. Using this method, one dimensional data (1d data) can be transformed into multi-dimensional data called attractors. Many phenomena in the real world are represented by attractors, which can be treated as mathematical models. Attractors are also applied to systems, physical phenomena, and economic phenomena in the living body [2]. In recent years, there is a lot of research on time series data analysis using neural networks (NN) such as 1-dimensional convolutional neural networks (1d-CNN) and recurrent neural networks (RNN). The advantage of NN is that they can capture features by themselves. However, even with these models, it is difficult to learn numerical and time-series features of time series data with very complex oscillations. Therefore, data preprocessing is important in NN. In this study, to solve this problem, this research firstly performs optimal noise reduction of biological signals. Next, the data is extended to 3D data using a time-delay coordinate system and the dimensional reduction methods to compress the data into 1D data. The objective is to find features that can be easily learned by NN from time series data with complex oscillations.

ORCID iDs Ryosuke Shimizu: 0000-0001-9684-7334, Yoko Uwate: 0000-0002-2992-8852, Yoshifumi Nishio: 0000-0002-0247-0001

2. Convolutional Neural Network

Convolutional Neural Network (CNN) is one of the NN and is mainly used in image recognition [3]. In 1982, Fukushima and Miyake proposed ”New Recognition TRON”, the predecessor network of CNN [4]. However, CNN are required high computational power. To compensate for these shortcomings, in 2015, Kiranyaz proposed the first compact and adaptive 1d-CNN that works directly with patient-specific ECG data [5]. Today, 1d-CNN have quickly achieved state-of-the-art performance levels in several applications such as biomedical data classification and early diagnosis, structural health monitoring, anomaly detection and identification in power electronics and electrical [6]-[8]. CNN is a model with a hierarchical structure that overlaps layers called convolutional layers, and the accuracy has been improved by deepening this hierarchical structure. In 1998, Le Cun Laba LeNet-5 has five layers. Categorizing handwritten characters using CNN. In 2012, eight-layer AlexNet won the ILSVRC image classification competition In the 2014 competition, VGG Net 19. GoogleNet 22 layers have been layered to further improve accuracy. In the 2015 competition, the Residual Network (ResNet) layered 152 layers and won the competition of the year. Now ResNet is popular as one of the standard CNN models [9]. In this study, 1d-ResNet is used for time series classification.

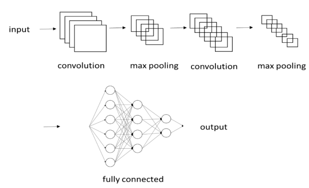
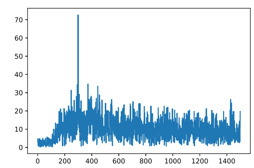


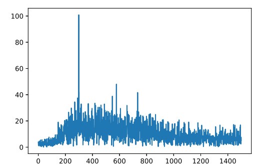
Figure 1: An example of CNN structure.

3. Dataset

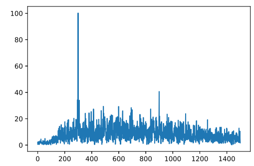
The UCR repository [10] data used in this study is the EMG (EMG) of hand movements. The participants were asked to repeat the following six movements, which are considered basic hand movements. Data (a) is the movement to hold a spherical tool. Data (b) is for holding a small tool. Data (c) is an action for grasping with the palm of the hand. Data (d) are movements for holding thin and flat objects. Data (e) is for holding a cylindrical tool. Data (f) is an action to support a heavy object. An example of the EMG time series data for these six movements is shown in (Fig. 2).



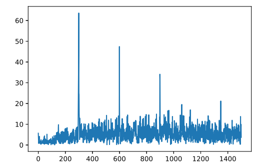
(a) hand movement



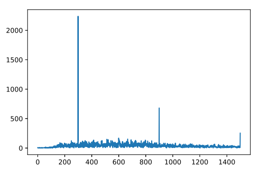
(b) hand movement



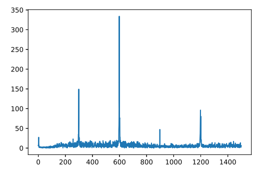
(c) hand movement



(d) hand movement



(e) hand movement



(f) hand movement

Figure 2: Six hand movements

4. Proposed method

In this study, we propose a method of feature translation for using multidimensional features of time series data. Figure 3 shows the flow of the proposed method of this study.

1. The data is smoothed for feature extraction.
2. Transformed into three dimensional attractor.
3. Compressed to extract the necessary features.

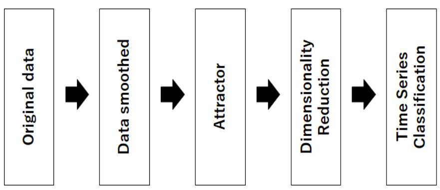


Figure 3: 1d-ResNet structure used in this study.

(a) The LOWESS smoothing method

The LOWESS smoothing method is a general method for determining smoothing lines. LOWESS stands for locally-weighted scatterplot smoother. The routine selects a fixed percentage of all points using the data that are closest neighbors in x-values on either side of the (x, y) point. A weighted linear regression is performed on each data point. This will give the point closest to each x-value the greatest weight in smoothing, limiting the effect of outliers. Parameters can be specified to change the degree of smoothing and influence from outliers. The weights of the smoothing parameters can also be specified. The larger the weight, the more the smoothed value follows the data; the smaller the weight, the smoother the pattern of smoothed values.

(b) Attractor construction

Time-delayed coordinate systems are commonly known as dimensional dilation methods by Takens ’ Embedding Theorem. It is a method of dilation of dimensions. Let the value of the data at a certain time be $x(n)$. Further, let the value of time delay be τ. This system is represented by Eq. (1) and is shown in Fig. 4 [13]. In this study, we extended the data to three dimensions.

$$f(x) = [x(n),\ x(n+\tau),\ x(n+2\tau)\ ...] \tag{1}$$

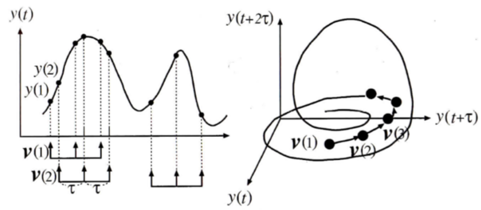


Figure 4: Time delay coordinate system.

(c) Dimensionality reduction

Dimensionality reduction refers to the reduction of high-dimensional feature vectors to low-dimensional vectors while retaining the distribution information of the original dataset. The procedure of Principal Component Analysis (PCA) method is shown in the following steps.

1. Determine the axis that can maximize the variance
2. Determine the axis of the second that is orthogonal to the first
3. Get the subspace as eigenvectors by singular value decomposition
4. Compute lower-dimensional embedding

PCA is one of the most widely used low-dimensional embedding methods. As the name implies, it is an algorithm that searches for the important element principal component in the observed data.

5. Simulation Model

In this study, the 1d-ResNet is used for classification model.

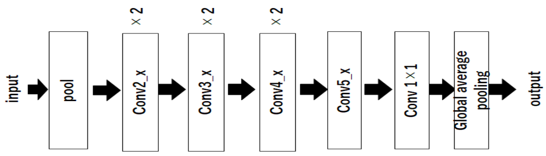


Figure 5: 1d-ResNet structure used in this study.

ResNet is commonly used for the classification of image data. ResNet is the winner network of ILSVRC 2015, and it is the network that enable to have deep layered networks without vanishing gradient problem by using residual learning that is shown in Fig. 5. By inserting the short-cut connection that is expressed by the longest arrow over 3 convolutional layers, previous informa- tion is learned, and it avoid vanishing gradient problem. In Fig. 6, Convolution (1 × 1, 128) indicates convolution layer with 1 × 1 filter and 128 channel.

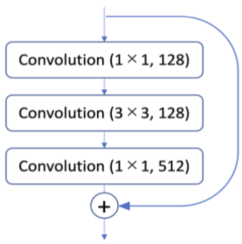


Figure 6: The structure of residual learning.

Table I shows the batch size, epochs, and learning rate in this study. Batch size is the amount of data processed at once. The epoch is the number of the learnings. Simulation terminates when val loss no longer decays during 1500 epochs. The learning rate is the degree of learning progress. By setting as shown in Table I, the training accuracies reach 100%.

Table 1: Learning parameters.

	Batch size	Learning rate	Epoch
1d-ResNet	42	0.0001	Earlystopping

6. Simulation results

We investigate the average of 5 times of test accuracy. 1d-ResNet is used to investigate the accuracy of 6-value classification. Table II shows the accuracy of NN with and without the smooth process of the proposed method was compared.

Table 2: Test accuracy of the proposed method.

	test accuracy
PCA(not smoothed)	0.836
PCA(p = 10)	0.819
PCA(p = 50)	0.861
PCA(p = 100)	0.840

The value p is the range parameter of the subset when performing weighted linear least squares regression in lowess regression. Optimizing this value improved the accuracy. The best accuracy was obtained when p = 50 in the lowess regression. When p = 10, the accuracy was lower than the accuracy when the smooth process was not used. Figure 8 shows the image of attractor without the smooth process. Figures 9-11 are images of the attractor composed of waveform (a) in Fig. 2 before dimensionality reduction. The attractors are characteristic, and it is thought that the smooth process has brought out the characteristics.

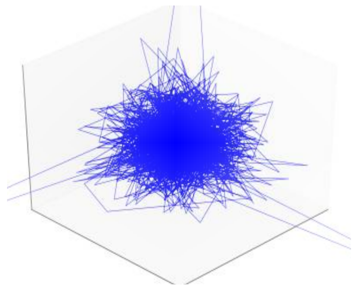


Figure 7: not smoothed Attractor.

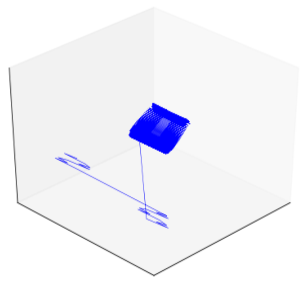


Figure 8: Attractor at p = 10.

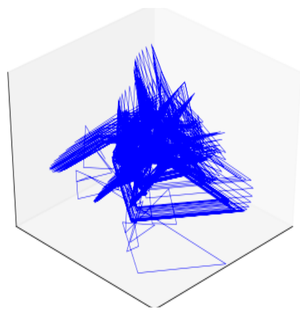


Figure 9: Attractor at p = 50.

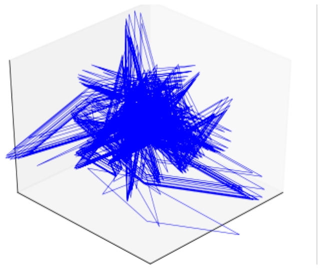


Figure 10: Attractor at p = 100.

7. Conclusion

In this study, we extended smoothed time series data to 3D space using time-delay coordinate method and compressed the data. we checked how the accuracy of time series classification by 1d-ResNet changes by optimizing the number of subset parameter during smoothing. The NN accuracy was also compared between the cases without the smoothing process and with the proposed method. The results show that the accuracy of the test Smoothing with appropriate values allowed the noisy data to perform feature extraction well. Image of the the attractors of the proposed method were found to be distinctive. We will use these results to investigate whether they are valid for other data as well. We will also use models such as RNN that are specialized for analyzing the flow of time series or construct optimal extended dimensions.

References

[1] F.Takens, Detectingstrange attractors intur bulance,LectureNotes in Mathematics, Vol.898(Springer-Verlag. 1981) 366-381.

[2] Christian Oestreicher,"A history of chaos theory", doi: 10.31887/DCNS.2007.9.3/coestreicher.

[3] M. Matsugu, K. Mori, Y. Mitari and Y. Kaneda, "Subject independent facial expression recognition with robust face detection using a convolutional neural network", Neural Networks 16., vol.5, pp.555-559, 2003.

[4] K. Fukushima, S. Miyake, Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position, Pattern Recognit. 15 (1982) 455469. doi 10.1016/0031-3203(82)90024-3.

[5] S. Kiranyaz, T. Ince, R. Hamila, M. Gabbouj, Convolutional Neural Networks for patient specific ECG classification, in: Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS, 2015. doi:10.1109/EMBC.2015.7318926.

[6] Shraddha R. Bajare, Vaishali V. Ingale, "ECG Based Biometric for Human Identification using Convolutional Neural Network", Computing Communication and Networking Technologies (ICCCNT) 2019 10th International Conference on, pp. 1-7, 2019.

[7] Pragya Sharma, Swet Chandan, B. P. Agrawal, "Vibration Signal-based Diagnosis of Defect Embedded in Outer Race of Ball Bearing using 1-D CNN", Computational Performance Evaluation (ComPE) 2020 International Conference on, pp. 531-536, 2020.

[8] Imene Mitiche, Alan Nesbitt, Stephen Conner, Philip Boreham, Gordon Morison, "1D-CNN based real-time fault detection system for power asset diagnostics", Generation Transmission Distribution IET, vol. 14, no. 24, pp. 5766-5773, 2020.

[9] He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. InProceedings of the IEEE. conference on computer vision and pattern recognition(pp. 770-778).

[10] Yanping Chen, Eamonn Keogh, Bing Hu, Nurjahan Begum, Anthony Bagnall, Abdullah Mueen and Gustavo Batista (2015). The UCR Time Series Classification Archive.

[11] E. F. Delagi, A. O. Perotto, J. Iazzetti, and D. Morrison, Anatomical guide for the Electromyographer: The Limbs and Trunk, Charles C Thomas Pub Ltd, 5ht ed., 2011.

[12] T. S. Saponas, D. S. Tan, D. Morris, R. Balakrishnan, J. Turner, and J A. Landay, Enabling Always-Available Input with Muscle-Computer Interfaces ´ in 22nd annual ACM symposium on User interface software and technology, Association for Computing Machinery, New York, USA, pp 167-176, 2009.

[13] K. He, X. Zhang, S. Ren and J. Sun, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification ", IEEE International Conference on Computer Vision and Pattern Recognition, arXiv:1502.01852, 6 Feb 2015.

[14] K. Aihara, T. Ikeguchi, Y. Yamada and M. Komuro, " Fundamentals and applications of chaotic time series analysis ", Industrial book, p.23, 2000